A Theoretical Framework for Integrating Peer Assessment in Teacher Education

Running Head: A THEORETICAL FRAMEWORK FOR INTEGRATING PEER ASSESSMENT IN TEACHER EDUCATION

This is a pre-print of the article that was published as:

Sluijsmans, D. M. A., & Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation, 32,* 6-22.

http://www.elsevier.com/wps/find/journaldescription.cws_home/497/description#description

A Theoretical Framework for Integrating Peer Assessment in Teacher Education

Dominique Sluijsmans, Frans Prins

Open University of the Netherlands, Heerlen, The Netherlands

Correspondence concerning this paper should be addressed to Dominique Sluijsmans, Open University of the Netherlands, Educational Technology Expertise Center, P.O. Box 2960 6401 DL Heerlen, The Netherlands, voice: ++31-45-5762178, fax: ++31-45-5762802, e-mail: dominique.sluijsmans@ou.nl

Keywords: Constructive Alignment; Peer assessment; Student Involvement; Teacher Education; Training

Abstract

Peer assessment can be a valuable learning tool in teacher education, because it supports student teachers to acquire skills that are essential in their professional working life. In this article a theoretical framework is presented in which the training of peer assessment skills by means of peer assessment tasks is integrated in teacher education courses. Theories about constructive alignment, student involvement, instructional design, and performance assessment underlie the framework. Furthermore, an overview of three empirical studies is provided to illustrate the implementation of the framework in a teacher training context. Results show that the framework offers powerful guidelines for the design and integration of peer assessment activities in teacher training courses. In general, the peer assessment tasks that were embedded in the courses lead to a general improvement in students' peer assessment skills as well as their task performance in the domain of the course. Implications for course and curriculum design are discussed.

A Theoretical Framework for Integrating Peer Assessment in Teacher Education

Teacher training colleges face the complex task to educate student teachers who, in turn, have to educate pupils in elementary schools in the future. Two recent trends in education, that is, the design of more *competency-based curricula* and the *involvement of students in assessment* (Verloop & Wubbels, 2000), urge teacher training colleges to modify their educational practices. In this article, we argue that the use of peer assessment in the curriculum of student teachers fits well in a competency-based curriculum and that it fosters student involvement in assessment. To be effective, however, peer assessment training should be embedded in the existing course material that is designed according to a performance-based approach (Mehrens, Popham, & Ryan, 1998). We present a theoretical framework for integrating peer assessment in teacher education as well as three empirical studies in which teacher training courses that were designed according to the framework were evaluated.

Peer assessment and competency-based education

Institutions of higher education in general are continuously challenged with a demand for competency-based learning. A curriculum should focus more on competencies such as learning to learn, interactive skills, communication skills, information processing, problem-solving, and reflective skills (Tillema, Kessels, & Meijers, 2000). Skill-based learning is an ongoing issue in the domain of teacher education (Darling-Hammond & Snyder, 2000; James, 2000; Kremer-Hayon & Tillema, 1999; Willems, Stakenborg, & Veugelers, 2000). A number of teacher training colleges collaboratively formulated a broad scale of skills student teachers need to develop. These skills of a primary school teacher are reported in a vocational training profile (LPC, 1995), which consists of 41 skills. These skills represent the overall accepted knowledge, proficiency and attitudes a primary school teacher needs to acquire.

The skill to assess the work of peers is a specific skill of the vocational training profile of primary school teachers. The process whereby individuals evaluate the performance of

their peer is called *peer assessment* (Falchikov, 1995; Freeman, 1995). In our view, peer assessment is a powerful didactical method for teaching skills that are important for the teaching domain for at least four reasons. First, teachers have to work together, learn from each other and become a member of a learning organisation (Verloop & Wubbels, 2000). Besides, the importance of communication between teachers in schools has been endorsed by many researchers (Cohen, 1994; Johnson, Johnson, & Johnson-Holubec, 1992; Sharan & Sharan, 1994; Slavin, 1995). In a peer assessment task, students have to communicate and collaborate and, thus, are able to acquire communication and collaboration skills. Second, discussion about reflection is an ongoing issue in teacher education (e.g., Korthagen, 1985, 2001; Newman, 1996; Reilly Freese, 1999; Richert, 1999). Encouraging students to assess each other's contributions to discussion and discourse, as in peer assessment, is further exposing them to the skills of critical reflection and analysis (Birenbaum, 1996; Sambell & McDowell, 1998). Reflection skills are conditional for making reliable judgments about peers' work. Thus, peer assessment fosters reflection and the development of reflection skills. Third, student teachers will become assessors in their own classroom and, therefore, they will have to design assessments as prospective teachers of children in primary schools. It is therefore advisable to teach student teachers how to make critical judgements about the performance of their peers, and, later on, about performances of children. The last reason for the importance of peer assessment in teacher education is that after students have left higher education, they are likely to rely heavily on the judgement of their peers to estimate how effective their performances in the school are (Brown, Rust, & Gibbs, 1994). Being able to interpret the work of colleagues and peers is a necessary prerequisite for professional development and for improving one's own functioning (Verloop & Wubbels, 2000). Training in peer assessment skills stimulates this mutual influence to take place at a professional level.

Performance assessment as a fundament for peer assessment tasks

Peer assessment is regarded as a learning tool that may have positive effects on skills that are relevant for teachers. In our view, performance *assessment* should be the fundament for peer assessment tasks. Performance assessments are described in terms of a certain performance that is content related and is perceived as worthwhile and relevant to the student in relation to their future profession. This performance may or may not represent an authentic situation (Wiggins, 1989). Performance assessment focuses on the ability to use combinations of acquired skills and knowledge, and therefore fits in well with the theory of constructive alignment and powerful learning environments (Linn, Baker, & Dunbar, 1991; Birenbaum, 2003).

Performance assessments require individuals to apply relevant knowledge and skills in context, not merely completing a task on cue. Students are observed while they are performing, products they create are examined, and the level of proficiency demonstrated is judged. Performance assessment can be based on multiple products or processes, for example essays, reflection papers, oral assessments, simulations, process-analyses, group-products, and work-samples. Judgments are made about the level of achievement attained by comparing student performance to predetermined standards. All students have the opportunity to attain the standards, whereby they can play a crucial role in making judgments about the performance of their peers and *defining appropriate criteria* for these performances. The importance of the negotiation about criteria has already been stressed in several studies (Boud, 1995; Orsmond, Merry, & Reiling, 1996; 1997; 2000). Or as Stiggins stated: “Once students internalise performance criteria and see how those criteria come into play in their own and each other’s performance, students often become better performers” (1991, p. 38).

As opposed to most traditional forms of testing, performance assessments do not provide clear-cut right or wrong answers. The performance is evaluated in a way that allows for informative scoring on multiple criteria. This is accomplished by creating assessment

forms. In these forms teachers determine at what level of proficiency a student is able to perform a task or display knowledge of a concept. For example, the different levels of proficiency for each criterion can be defined. Using the information of the assessment form, feedback is given on a student's performance either in the form of a narrative report or a grade. A criterion-referenced qualitative approach is desirable, whereby the assessment will be carried out against the previously specified performance criteria. An analytic or holistic judgment then is given on the basis of the standard the student has achieved on each of the criteria. The basis of the effective application of performance assessment methodology is thoroughly trained raters relying on sound performance criteria to observe and evaluate student responses to quality exercises (Stiggins, 1994).

Designing performance assessments

A common error in designing a course or unit of study is to leave the development of the performance assessment as a final activity (Airasian, 1991). The compatibility between learning, instruction and assessment is a basic assumption for our framework. Biggs' (1996, 1999, 2001) theory of constructive alignment and Stiggins' (1987) approach are useful to design courses and performance assessments. Four steps can be taken to design courses in which instruction and assessments are completely aligned. First, teachers must have a clearly defined purpose of a course. The concepts, skills, and knowledge that have to be assessed as well as the level at which students should be performing must be determined (Stiggins, 1987). Second, it must be decided what type of activity best suits the assessment needs. This can result in a skill decomposition in which the relevant skills are ordered hierarchal, or in which they are organized in a concept map. In the third step, decisions should be made concerning the assessment task. Issues that must be taken into account are time constraints, availability of resources, and how much data is necessary in order to make an informed decision about the quality of a student's performance. Finally, after the assessment task is determined, the

elements of the task that determine the measure of success of the student's performance needs to be defined. Sometimes, these can be found in so-called job-profiles.

Most of the time, teachers have to analyse skills or products to identify performance criteria upon which to judge achievement, which is not an easy task. Criteria should be significant, specifying important performance components, represent standards that would apply naturally to determine the quality of performance when it typically occurs (Quellmalz, 1991). The criteria must be communicated clearly to and be able to be understood by all involved. Communicating information about performance criteria provides a basis for the improvement of that performance. When a teacher has passed through this procedure, study tasks can be designed in which students are prepared for the performance assessment. These study tasks are directly related to the performance assessment task at the end of the course.

Designing courses in which peer assessment is integrated

According to Sluijsmans, Dochy, and Moerkerke (1999), teacher educators should be supported in the design of learning activities in which peer assessment is integrated. The abovementioned design guidelines of Stiggins (1987) are helpful. Step 1 of the design process is to define the purpose of a course. What should be emphasized is that a course that includes peer assessment tasks contains multiple learning goals. The performance of the student at the end of the course is content related and can be labelled as the *first order goal* of a course. Acquiring peer assessment skills is subsequently integrated as a *higher order goal* in a particular course. Students learn to evaluate the course-content related performances of peers at the end of a course. Peer assessment can thus be considered as a performance assessment that is superposed on the content-related performance assessment. When the acquisition of peer assessment skills is one of the purposes of a course, at the end of the course students should be capable of making arrangements in which they negotiate with students of similar status about the design and appropriate criteria of specific study tasks and performances. The

student should also be able to take the responsibility to make critical judgements about the performances of a peer applying the appropriate criteria. It should be noted that peer assessment skills are not easily and automatically acquired. Peer assessment is considered as a complex skill that needs to be developed (Birenbaum, 1996; Reilly Freese, 1999; Sluijsmans, Dochy, Moerkerke, & Van Merriënboer, 2001). Students who are novices in assessing are insecure about their ability to assess and indicate that they need more guidance on the marking criteria (Cheng & Warren, 1997; Woolhouse, 1999). Normally students need explicit training in assessment techniques during the course to make reliable and acceptable assessment reports (Boud, 1990; Hanrahan & Isaacs, 2001).

The method of skill decomposition is applied to identify constituent skills (Van Merriënboer, 1997). The task to peer assess is broken down into separate skills and these skills are practiced one at a time, before being recombined and practiced as a complete task (step 2). In Figure 1 the skill of peer assessment is modelled. Each constituent skill of the peer assessment is further described (see Table 1). Data for this decomposition were gathered through literature review and feedback from experts in the area of peer assessment. The horizontal relationship in Figure 1 illustrates which more specific skills are necessary in order to be able to perform the skill under consideration. The vertical relationship illustrates which other skills are necessary to be able to perform the peer assessment skill.

****INSERT FIGURE 1 ABOUT HERE****

****INSERT TABLE 1 ABOUT HERE****

The performance assessment task for determining the quality of the peer assessment skill should then be chosen (step 3). Normally this task is to write an assessment report about the performance of a peer at the end of the course. This assessment report can be used for summative assessment purposes, while the embedded peer assessment tasks have a more supportive function in developing the skills that are conditional for conducting a peer

assessment. Both the quality of the assessment reports and the performance assessments can be examined by the teacher educator. Assessing the peer assessment skill is however still very rare in teacher education.

Based on the skills presented in the model, criteria have to be defined for a good assessment report (step 4). Written assessments of expert assessors can be used to determine these criteria. Criteria are determined regarding the use of adequate criteria, giving feedback and the style of a written assessment report. In practice, students write a qualitative assessment report about a performance of one or more peers on a blank peer assessment form. A rating form has to be developed to analyse the quality of the peer assessments that were written by the students. Naturally, this rating form is based on the criteria for a good assessment report. Teacher educators use the rating form to determine the quality of the assessment skill.

An Integrated Framework For Training Assessment Skills

In Figure 2, it is illustrated how the concepts presented in the previous sections are integrated in a framework that underlies our three empirical studies.

****INSERT FIGURE 2 ABOUT HERE****

Overviewing the concepts discussed in the previous sections, it can be concluded that there are two parallel paths, illustrated by the shaded arrows. In the ‘first-order course design path’, students are guided in the acquisition of content related skills through study tasks with the aim to meet the criteria for the content-based performance assessment. The second path is the ‘higher-order course design path’, in which students are supported in the acquisition of peer assessment skills, by means of peer assessment tasks (PA-tasks). These peer assessment tasks, which are superposed on the regular study tasks, are characterised by collaborative learning, more specific by social interaction, individual accountability and positive

interdependence (Slavin, 1989). Students work towards two assessments: a content-related assessment (*the first order course goal*) and a peer assessment (*the higher order course goal*).

The two paths are integrated (see the two dotted arrows), in other words, the peer assessment tasks are completely embedded in the study tasks of the course, because the content of the study tasks provide input for the peer assessment tasks. The first-order and higher-order course design are the basic elements of the framework, and are defined from the theory of student involvement, the constructive alignment theory and the design principles of Stiggins (1987). At the end of a course, students have to carry out a performance assessment, which is subsequently object of the peer assessment.

Empirical studies

Three experimental studies were conducted within the context of teacher education to illustrate how peer assessment can be integrated in a course according to the theoretical framework presented in Figure 2. Moreover, in these studies, the effects of an embedded training in peer assessment skills on students' performance in their peer assessment skills and content-based skills were examined. The following research questions were explored:

1) Does training in peer assessment lead to the development of the skill to assess the work of peers (the higher order goal)?
2) Does following a training in peer assessment lead to an improved task performance in the domain of a course (the first order goal)?
3) What are perceptions of students and teachers regarding the implementation of the framework?

We expected that the training in peer assessment had positive effects on the development of peer assessment skills as well as on task performance in the domain of the course.

In the following sections we first describe the method that was directive for each study. Then we present the design and procedure of each study. Overall results of the studies will be summarised.

Method

Participants

The sample in each study consisted of first-year and second-year students of a Primary Teacher Training College in the Netherlands.

Materials

Peer assessment form. At the end of the selected courses in each study, students had to assess the products that were object for performance assessment on a blank peer assessment form.

Rating form. To analyse the quality of the peer assessments that were written by the students, a rating form with underlying variables derived form the peer assessment model was developed. Detailed information about the rating forms can be found in Sluijsmans, Brand-Gruwel, Van Merriënboer (2002) and Sluijsmans, Brand-Gruwel and Van Merriënboer & Bastiaens (2003). Variables were related to the skills of defining criteria, giving feedback, and writing an assessment report. In each study independent research assistants scored the peer assessment forms with the rating form. For each variable the interrater-reliabilities were calculated. These reliabilities were acceptable for all variables in each study (Cohen's Kappa >.95).

Examinations. To measure an effect of the peer assessment training on the performance of students, the marks on the performance assessments given by the teacher were analysed. These performance assessments were in the studies a lesson plan for discovery learning (study I), a video on creative learning (study II) and a reflection paper (study III).

Student questionnaire and structured student interviews. Before and after the courses in each study, students filled out a questionnaire about their perceptions on instruction and assessment. Items were divided among several variables related to instruction, vision on instruction and assessment and the role of the student in assessment. The pre-test was carried out to investigate the students' perceptions on prior courses that were comparable to the courses that were selected in the studies. These prior courses were not designed in a skill-based way. The post-test concerned students' perceptions after the redesigned course according the approach chosen in the framework. The students had to score the items on a five-point Likert scale, varying from 'I totally disagree' to 'I totally agree'.

Teacher questionnaire and interview. The teacher educators who were involved in the course in the first and third study evaluated the four peer assessment tasks by means of a short questionnaire. The questions concerned issues related to the framework on the basis of which the courses were redesigned and the peer assessment tasks were integrated. The questions were related to two phases, the design phase of the course and the implementation phase. Regarding the design phase, questions were asked about their experiences with the redesign of the course and their co-operation with the other colleagues. Questions related to the implementation phase concerned the experiences with the instruction of the peer assessment tasks and their vision on assessment and instruction and the role of students and themselves.

Design and procedure study I

The first study was carried out with 93 second-year student teachers. A second-year course on discovery learning was selected for redesign. The former version of the course was designed from the perspective of the content domain. A problem of this course was that students felt that discovery learning was basically linked to the physics domain, although four other domains were also involved. Another problem was that students worked on several course objectives that led to a high workload, without thoughtful consideration why they had

to work on specifically those products. To solve these problems, the existing course was redesigned from a skill-based perspective for the purposes of the present study. It was decided that the new course objective was that students were trained in their skill to design a lesson plan on discovery learning in the context of one of the five content domains. In operational terms, at the end of the course students had to deliver a lesson plan that was related to one of the five content domains. Therefore, the 93 student teachers were randomly distributed amongst the pedagogy domain (n = 20), the physics domain (n = 21), the philosophy domain (n = 21), the mathematics domain (n = 21), and the music domain (n = 10). Before the design of the concrete study tasks, the involved teachers decomposed the skill of designing a lesson plan on discovery learning similar to the way the skill to assess was analysed (Van Merriënboer, 1997). This resulted in four main sub skills students had to acquire with regard to the design of a lesson plan for discovery learning: 1) introducing a problem in a classroom with pupils; 2) posing the right questions to the pupils in relation to the introduced problem; 3) analysing the problem with pupils, and 4) solving the problem with pupils. A study task was designed for each of the four skills in each of the five content domains. The whole course enclosed six classes of an hour and a half each in a period of four weeks: an introductory class, four regular course classes, and one class in which the students assessed the end product of peers. In the four regular classes, the content related study tasks regarding discovery learning were instructed, based on the four skills. A complete overview of the organisation of the course is given in Figure 3.

****INSERT FIGURE 3 ABOUT HERE****

Because the peer assessment skill is too complex to be trained in only one course (Van Merriënboer, 1997), for this study it was decided to train the students in the first main constituent skill: *defining criteria*. Half of the group was trained in peer assessment skills (experimental groups) and the other half was not (control groups). Before the start of the

course, the students filled out a student questionnaire as a pre-test. Both the control groups and the experimental groups attended the regular classes as presented in Figure 3. The experimental groups followed four embedded peer assessment tasks. In these tasks, that were embedded in the four regular course classes, students had to define measurable criteria that were related to each of the four skills for designing a discovery learning lesson plan. For this, the teacher presented examples of valid and invalid criteria. Each peer assessment task was characterised by interactive discussions between the students to foster collaborative learning and paid attention to the skills that are related to defining criteria. Students were encouraged to think about 'personal' course objectives and the relation between course objectives and the study tasks.

The time students in the control groups spent on the regular classes was the same the students in the experimental groups spent on the classes and the peer assessment tasks together. Thus, the students in the control groups had relatively more time to discuss the content of the regular classes, because they did not receive the peer assessment training. In each peer assessment task, a part of the whole criteria list for a lesson plan was developed. This was done through constructive discussions guided by the teacher. The students were encouraged by the teacher to make their personal ideas explicit. At the end of the fourth and last peer assessment task, the students had a list of ten criteria. During the course, all students worked in dyads on the end product, which was a design of a lesson plan for an elementary school. At the end of the course the dyads had to present their end product to the rest of their group. In the last class of the course, both the students of the control groups and the experimental groups were instructed to write a qualitative peer assessment with regard to the content of the lesson plan of the peer dyads. Each student wrote four peer assessments, because in each group there were four other dyads to assess. After the course, all students filled out the same questionnaire as in the pre-test. The teachers who taught the experimental

groups filled out the teacher questionnaire after each peer assessment task. In the two weeks after the course, the teachers and 16 students were interviewed.

Design and Procedure Study II

In the second study a similar experiment was set up, but in this study students were trained in several assessment skills, instead of only the skill of defining criteria. For the purpose of this study, a second-year course on creative learning was chosen. The teachers that were jointly responsible for this course first redefined the course objective because the course objectives were not revised for several years and teachers had developed multiple perspectives on what the content should be. It was decided that students were guided in the content skill “designing a creative lesson”. At the end of the course, students had to make a videotape of a creative lesson that was designed and carried out by themselves. The four teachers collaboratively decomposed the skill of designing a creative lesson. This resulted in a concept map with a number of constituent skills. For the domains art, Dutch language, and music four one hour study tasks were defined, based on the constituent skills. In these tasks, students learned how each domain was related to creative learning and the design of creative lessons. The pedagogy teacher designed four one hour study tasks that integrated the tasks of the domains art, Dutch language, and music. The whole course enclosed an introductory class, sixteen study tasks (four tasks per domain), and a concluding class in which the peer assessment was organized.

Ninety-three student teachers were randomly assigned to control groups and experimental groups. The experimental groups were trained in three important assessment skills, namely defining performance criteria, giving feedback and writing assessment reports. Before the start of the course, all students filled out the questionnaire. During the course, all students worked in subgroups of five or six students on their design of a creative lesson and the group report. They prepared their lesson that was taped on video and subject of the peer

assessment. In between classes, each student worked individually on the individual report and the content domain related assignments. During the course the students of the experimental groups performed the four peer assessment tasks of one hour each. These tasks were embedded in the study tasks of the pedagogy domain, and were closely related to the study tasks concerning designing creative lessons. The training focused on the three main constituent skills of the peer assessment model (see Figure 1).

In *Task 1*, students were introduced to the meaning of peer assessment and the product that they were going to peer assess at the end of the course. This product was a video of a creative learning lesson taught by two second-year students. After this introduction students watched a creative learning lesson on video, discussed and elaborated on the fragments in which creativity was applied. This resulted in a first rough draft of the criteria that are required for a creative lesson.

In *Task 2*, the skill '*defining criteria'* was addressed. Examples of valid and invalid criteria were presented. Students then further elaborated on the rough criteria for designing a creative lesson they formulated in the first task. This exercise resulted in a list of 15 criteria that are required for a creative lesson, which were accepted by the students and the teacher.

Discussing the purpose and guidelines for giving constructive feedback, was the central topic in *Task 3*. In the peer assessment model, this is the skill '*provide feedback for future learning'*. First, the teacher asked the students what their ideas were about feedback and criticism. After a short discussion, the teacher presented an expert-assessment report to the students. This was an assessment report on the video lesson that was analysed in Task 1, which was written by two experts on creative learning. Students discussed the good examples of constructive feedback. At the end of the task, students had to give each other feedback on some aspects of their own work. The output of this task was a list of criteria for constructive feedback.

In *Task 4*, the students were trained in the third main skill of the peer assessment model, namely '*judge the performance of a peer'*. In this final task, the three prior tasks were integrated. To confront the students with ways in which an assessment report can be written, they analysed the expert-assessment report and discussed the structure that was applied by the experts. They also discussed the language used in the assessment, for example the use of naive words, like 'nice'. Based on the findings, students defined a peer assessment form.

Instead of these tasks the students of the control groups attended four extra hours in the pedagogic domain. During these hours, the control groups had the opportunity to elaborate on certain aspects of creative learning. At the end of the course, a peer assessment session was organized for each group (approximately 25 students), in which the video lessons of each subgroup were shown (four video lessons in each group). The peers were instructed to write a qualitative peer assessment report with regard to the content of the video lesson of each group. The experimental groups were free to use the output of the peer assessment tasks. For the peer assessment, the students from the control group had to use the regular course materials from the study tasks. Each student wrote three peer assessment reports, because in each group there were three other subgroups to assess. After the course, all students filled out the same questionnaire as in the pre-test.

Design and Procedure Study III

The findings of the first and second study set the design and goals of the third experiment. In this study, 110 first-year student teachers were longitudinally trained in peer assessment skills within three courses on mathematics. The study was set up according to a within subject repeated measures design. Students participated for a period of seven months in the experiment. In a two-hour intake session that took place a day before the start of the first mathematics course, the students carried out three activities: filling out the questionnaire, writing a reflection paper about prior experiences in mathematics, and assessing an

anonymous reflection paper. This anonymous reflection paper was previously marked as an 'unsatisfactory' one by the mathematics teacher. After the intake, all students attended three successive courses on mathematics. Within the three courses students were confronted with basic skills that are required for teaching mathematics to pupils. Besides that, the students had to write a reflection paper after the first course, which could be improved after the second and third course to submit the final version of the paper two weeks after the last feedback session. All students received training in the assessment skill during the courses. The assessment training was directed at three topics: what are important criteria for a reflection paper (four tasks in the first course), how to give feedback (two tasks in the second course), and how to write an assessment report (two tasks in the third course). In this third course for example students developed a peer assessment form based on an expert assessment report that was written by the mathematics teacher. The output of the first part of the training was a list of 19 criteria for a reflection paper. Students agreed in negotiation with the mathematics teacher that a good reflection paper contains for example self-criticism, work field experiences, personal expectations, and strengths/weaknesses.

In the second training, integrated in the second course, students developed guidelines for giving feedback. One guideline that students agreed on was that it would be positive for a peer to mention their own learning experiences in the assessment report.

In the third and last part of the training, which was embedded in the third course, students worked on a peer assessment form and decided what is important in the writing of an assessment report. An expert assessment report acted as an example.

Students were instructed that the criteria, feedback rules and structure guidelines, derived from the peer assessment training, could be helpful in writing the reflection papers and the peer assessment. After each course, the students had to send their reflection paper to the other students. This was done using the facilities of Blackboard®, a virtual learning

environment. Each student had to assess the reflection paper of another student, which was organized in a way that every student had to assess and was assessed by different peers. After each course, a feedback session was organized, chaired by the mathematics teacher. In these sessions, in which a group of ten to twelve students participated, each student had to present orally his or her assessment report. The written report was given to the assessed student after the feedback session. The students used the feedback of the peers to rewrite and improve their reflection paper. The student feedback can be regarded as the formative assessment of the papers. To decrease the test anxiety and to lengthen the period in which the peer assessment skills were trained, students received no grades of the mathematics teacher for their reflection paper after each course. The role of the teacher was limited to coaching and chairing in the feedback sessions. The reflection paper that was written based on the given peer feedback after each course was used for the final grade given by the mathematics teacher. After the third feedback session, an outtake session took place, similar to the intake. In this session, all students filled out the student questionnaire again. They also wrote an assessment report of the same reflection paper that was presented in the intake session.

General results

Three studies were set up to investigate the implementation and the effects of the framework, depicted in Figure 2. We will report the effects for each of the three research question separately. Table 2 shows the effects sizes for the variables that concern our research questions 1 (effect on the assessment skill) and 2 (effect on the content related skill). Cohen (1988) defined an effect size of approximately 0.2 as small, of 0.5 as medium, and of 0.8 as large.

Results showed positive effects of peer assessment training on the students' skill to assess the work of peers (research question 1). In the first study, it was found that the student teachers from the experimental groups were more capable in using the set criteria determined

during the peer assessment tasks than the student teachers of the control groups. Effects on other variables of the rating form were not found, probably because the training only focused on the skill ‘defining criteria’. In the subsequent study, the analyses of the qualitative peer assessment reports revealed that the experimental groups were more likely to use the criteria and to give more constructive comments than the student teachers from the control groups. The students who received training also scored higher on structure and used less naive words. In spite of the positive results reported in the first two studies, it was concluded that student teachers could not be regarded as expert assessors after a peer assessment training in one course. The training in the longitudinal study was integrated in three successive mathematics courses. Analysis of the peer assessments from the intake and outtake data revealed significant progress for most variables. All students used the criteria more adequately, gave more constructive feedback, and wrote more structured assessment reports after the training period of ten months. Students also adopted a more critical attitude in the outtake than in the intake.

The second research question focused on the effect of training peer assessment skills on students’ content-related performance. No difference between the performance quality of the students from the control and from the experimental group was found in study I. Explanations can be sought in the small progress in the peer assessment skill and the short training period. In study II, in which the whole peer assessment skill was trained, a positive effect of the peer assessment training on the actual learning results was found. The student teachers from the experimental groups outperformed the students from the control groups. This same result was found in the third study, where the total group of students wrote better reflection reports after the training than before the training.

We also examined the perceptions of the students and teachers regarding the implementation of the framework for integrating peer assessment in teacher education.

Results of the three studies showed that the students were more positive about the instruction and the integration of assessment and instruction after they took the redesigned course. The renewed course, which was designed from a skill-based perspective and consisted of tasks that fostered collaborative learning and interaction, led to an active participation of student teachers and the teachers. It can be concluded that the student teachers positively changed their view on aspects of learning and assessment. They were more satisfied about the classes and the criteria, and goals were more clear. The role of the teacher was also evaluated in a more positive way. The student teachers indicated that they are more capable in assessing than before the redesign of the course.

Teacher experiences were investigated in the first and third study. It appeared that the redesign process and the implementation phase demanded a lot of effort of the teachers that were involved in the courses. The need for revision of the courses did lead to some resistance. Some teachers doubted the value of the peer assessment and were sometimes reluctant to give up some part of their content expertise on behalf of the 'higher order' skills. In both studies, however, the teachers had no major problems in instructing the peer assessment tasks. The teachers indicated that implementing the peer assessment training led to a rethinking of the existing course and stimulated them to view the content from a different perspective.

Discussion

A theoretical framework for integrating peer assessment in teacher education was presented and evaluated in three studies. Each study was conducted within a teacher training context, in which the skill to assess peers' work is considered to be important. Our hypothesis was that if student teachers were trained to assess the performance of peers, this should lead to a general improvement in their peer assessment skills as well as their task performance in the domain of the course. Results of the studies corroborated this hypothesis. Nevertheless, it became apparent that the training had to be more systematic and of longer duration than was

feasible to organise in the available context and time span. The studies focused on short-term effects of the training in peer assessment. It is conceivable that peer assessment training and more critical reflection about assessment might have a long-term effect for students.

A relevant question for future research is how the design of courses and the design of assessment training is most conducive to skill acquisition. A reconsideration of the peer assessment model and the collaborative activities that were used in the framework appear to be desirable. It is also interesting to elaborate further on the relationship between peer assessment skill acquisition and content skill acquisition.

A restriction of the rating form used in the studies is that it measures the use of the appropriate criteria, and the extent to which students made positive, negative or constructive comments. This, however, does not necessarily mean that the students apply the criteria adequately and correctly. In-depth analyses of students' written assessment reports by content experts are recommended with regard to the limitations of the rating form. It needs further analysis and research to develop a reliable assessment instrument for analysing assessment skills.

By involving students in the design of instruction and assessment, they become aware of how and on what knowledge and skills they are assessed. Peer assessment can be conceived as an evaluative device, but in our approach it is also a powerful learning activity. The student is introduced as an important collaborator with the teacher in the creation of tasks as well as in developing guidelines for scoring and interpretation. Until today, many tests are kept under lock and key so students do not have knowledge about them ahead of time. By doing this, students will study in a particular way in the hope that this will improve their test performance, but there is virtually no way that students can 'learn by doing' in the way that they learn while engaging in a performance based assessment in which they were involved as one of the assessors (Frederiksen, 1984).

The framework has implications for course design. Within the framework of skill-based curriculum design, the educational material is no longer defined from the perspective of the content domain, but from the perspective of the skills (Tillema, Kessels, & Meijers, 2000). This means that skills are trained in the context of different content domains. Working with the framework encouraged teacher educators to think about the performance assessment at the beginning of a course design process. Assessment drives the learning process and overrides practically every other aspect of curriculum design (Longhurst & Norton, 1997). Changing assessment practices towards more performance-based approaches, will inevitably lead to a revision of instruction. Instruction, assessment, and learning and teaching strategies have to be completely aligned. Educators must develop appropriate assessments that have no single right answer and in which students' argumentation is key in defending their solution. The involvement of students in these processes implies an extra investment. Although the studies focused mainly on the training of student teachers, it became increasingly apparent that much effort has to be put into the professional development of teacher educators. Meanwhile, initiatives are conducted to define a vocational profile for teacher educators (Koster & Korthagen, 2001). The competencies of teacher educators are operationalised (Plake, Impara, Fager, 1993). Designing rich, authentic performance assessment is one of these competencies that deserves special attention. After all, assessment is the tail that wags the dog.

Changing assessment practices and views on learning and the role of students in this, is a considerable challenge in teacher education and higher education in general. The success of sound assessment practices lies on the one hand in a close relationship between learning, instruction, and assessment, on the other hand in qualified (student) assessors. The presented framework and the studies in this article attempted to make a contribution on both aspects. Important guidelines for practice are that students need to be guided in their skill-development, that a clear definition of performance criteria is crucial for effective

assessments, that collaborative activities need to be stimulated, and that teacher educators receive training in instructional design and alternative assessment approaches. From the 'practice as you preach' - philosophy, an important condition for successful initiatives on the student level is that teachers are receptive for self-reflection and change.

References

Airasian, P.W. (1991). *Classroom assessment*. New York : McGraw-Hill.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364.

Biggs, J.B. (1999). *What the student does: Teaching for quality learning at university*. Buckingham: Open University Press.

Biggs, J.B. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *Higher Education, 14*, 221-238.

Birenbaum, M. (1996). Assessment 2000: Towards a Pluralistic Approach to Assessment. In M. Birenbaum, & F. Dochy (Eds.), *Alternatives in Assessment of Achievements, Learning Processes and Prior Knowledge* (pp. 3-29). Boston: Kluwer Academic Press.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), Optimising new modes of assessment: In search of qualities and standards (pp. 13-36). Dordrecht: Kluwer Academic Publishers.

Boud, D. (1990). Assessment and the promotion of academic values. Studies in *Higher Education, 15,* 101–111.

Boud, D. (1995). Enhancing learning through self-assessment. London: Kogan Page.

Brown S., Rust C., & Gibbs G. (1994). *Strategies for diversifying assessment.* Oxford Centre for Staff Development, Oxford.

Cheng, W., & Warren, M. (1997). Having second thoughts: student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22,* 233–239.

Cohen, E. G. (1994). Restructuring the classroom: conditions for productive small groups. *Review of Educational Research, 64,* 1-35.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16,* 523-545.

Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International, 32,* 175–187.

Frederiksen, N. (1984). The real test bias: influences of testing on teaching and learning. *American Psychologist, 3,* 193-202.

Freeman, M. (1995). Peer assessment by groups of group work *Assessment and Evaluation in Higher Education, 20,* 289–300.

Hanrahan, S., & Isaacs, G. (2001). Assessing self- and peer assessment: the students' views. *Higher Education Research and Development, 20,* 1, 53-70.

James, P. (2000). A blueprint for skills assessment in higher education. *Assessment and Evaluation in Higher Education, 25,* 353-367.

Johnson, D. W., Johnson, R. T., & Johnson-Holubec, E. (1992). *Advanced cooperative learning.* Edina: Interaction Book Company.

Korthagen, F.A.J. (1985). Reflective teaching and pre-service education in the Netherlands. *Journal of Teacher Education, 36,* 11-15.

Korthagen, F.A.J. (2001). *Linking practice and theory: the pedagogy of realistic teacher education.* Mahwah, NJ: Lawrence Erlbaum Associates.

Koster, B., & Korthagen, F. (2001). Training teacher educators for the realistic approach. In F. Korthagen (Ed), *Linking practice and theory: the pedagogy of realistic teacher education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kremer-Hayon, L., & Tillema, H.H. (1999). Self-regulated learning in the context of teacher education. *Teaching and Teacher Education, 15*, 507-522.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*, 5-21.

Longhurst, N., & Norton, L.S. (1997). Self-assessment in coursework essays. *Studies in Educational Evaluation, 23*, 319–330.

LPC (1995). *Beroep in beweging. Beroepsprofiel leraar primair onderwijs* [Profession in action. Vocational training profile for the primary school teacher]. Utrecht: Forum Vitaal Leraarschap.

Mehrens, W.A., Popham, W.J., & Ryan, J.M. (1998). How to prepare students for performance assessments. *Educational Measurement: Issues and Practice, 17*, 18-22.

Newman, S.J. (1996). Reflection and teacher education. *Journal of Education of Teaching, 22*, 297-310.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment and Evaluation in Higher Education, 21*, 239–249.

Orsmond, P., Merry, S., & Reiling, K. (1997). A study in self-assessment: tutor and students' perceptions of performance criteria. *Assessment and Evaluation in Higher Education, 22*, 357–369.

Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 25*, 23–38.

Plake, B.S., Impara, J.C., & Fager, J.J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement, Issues and Practice, 12*, 10-12.

Quellmalz, E.S. (1991). Developing criteria for performance assessments: The missing link. *Applied Measurement in Education, 4,* 319-332.

Reilly Freese, A. (1999). The role of reflection on preservice teachers' development in the context of a professional development school. *Teaching and Teacher Education, 15,* 895-909.

Richert, A.E. (1999). Teaching teachers to reflect: a consideration of programme structure. *Journal of Curriculum Studies, 22,* 509-527.

Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education, 23,* 391-402.

Sharan, Y., & Sharan, S. (1994). Group investigation in the cooperative classroom. In S. Sharan (Ed.), *Handbook of cooperative learning methods* (pp. 97-114). Westport: Praeger.

Slavin, R. E. (1989). Research on cooperative learning: An international perspective. *Scandinavian Journal of Educational Research, 33,* 231-243.

Slavin, R. E. (1995). *Cooperative learning: theory, research and practice*. Boston: Allyn & Bacon.

Sluijsmans, D.M.A., Brand-Gruwel, Van Merriënboer, J. (2002). Peer assessment training in teacher education. *Assessment and Evaluation in Higher Education, 27,* 5, 443-454.

Sluijsmans, D.M.A., Brand-Gruwel, S., Van Merriënboer, J., & Bastiaens, T, R. (2003). The training of peer assessment skills to promote the development of self-assessment skills in teacher education. *Studies in Educational Evaluation, 29,* 1, 23-42.

Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self- peer- and co-assessment. *Learning Environments Research, 1,* 293-319.

Sluijsmans, D., Moerkerke, G., Dochy, F., & Van Merriënboer, J.J.G. (2001). Peer assessment in problem based learning. *Studies in Educational Evaluation, 27,* 153-173.

Stiggins, R. (1987). Design and development of performance assessment. *Educational Measurement: Issues and Practice, 6,* 33-42.

Stiggins, R. (1991). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice, 10,* 7-12.

Stiggins, R.J. (1994). *Student-centered classroom assessment.* Columbus, OH: Macmillan.

Tillema, H.H., Kessels, J.W.M., & Meijers, F. (2000). Competencies as building blocks for integrating assessment with instruction in vocational education: a case from the Netherlands. *Assessment and Evaluation in Higher Education, 25,* 265-278.

Van Merriënboer, J.J.G. (1997). *Training complex cognitive skills.* Englewood Cliffs, NJ: Educational Technology Publications.

Verloop, N., & Wubbels, T. (2000). Some major developments in teacher education in the Netherlands and their relationship with international trends. In G. M. Willems, J. H. J. Stakenborg, & W. Veugelers (Eds.), *Trends in Teacher Education* (pp. 19-32). Leuven-Apeldoorn: Garant.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46,* 41-47.

Willems, G.M., Stakenborg, J.H.J., & Veugelers, W. (Eds.). (2000). *Trends in Teacher Education*. Leuven-Apeldoorn: Garant.

Woolhouse, M. (1999). Peer assessment: the participants' perception of two activities on a further education teacher education course. *Journal of Further and Higher Education, 23,* 211-219.

Table 1.

Description of the constituent peer assessment skills

First level	*Description*
Define assessment criteria	The student actively participates in a group discussion to reach a common understanding about the assessment criteria for the product to be assessed
Judge the performance of a peer	The student assesses individually a product of a peer by first analysing the product and then formulating the discrepancies between the product and the criteria. The formulated discrepancies are written down in a peer assessment report
Provide (anonymous) feedback for future learning	The student writes a feedback report that provides feedback for future courses. This feedback: • confirms that the peer's understanding of what the product required was correct; • helps the student to add information to his own knowledge when they experience an information gap; • helps the peer to replace the erroneous information with more accurate information.

Second level	*Description*
Develop 'personal' course objectives on the basis of given course objectives and group discussion	The student presents his personal interpretations of the course objectives and argumentates his view in a group session
Describe a personal report on course objectives	The student individually writes a report that reflects his interpretation of the course objectives
Couple course objectives to study tasks	In collaboration with his peers, the student relates the defined course objectives to the different tasks he has to carry out to reach the course objectives and formulates which part of the task contributes to which course objective
Develop measurable criteria for each study task	In collaboration with his peers, the student lists the criteria that were decided for the task; these criteria are the result of the task analysis
Analyse the performance of a peer	The student individually applies the assessment criteria to the product of the peer after reading the product and marks the evidence for the presence of the criteria
Formulate discrepancies in a peer assessment report	The student writes an assessment report on the quality of the product which reflects evidence for reaching the desired criteria at a certain level
Formulate points for improvement	The student writes individually a number of points for improvement based on the assessment criteria and the group discussions in which the assessment criteria were decided
Reflect on points of improvement for the peer	Based on the assessed product, the student individually presents and argumentates points for improvement to the peer

Third level	*Description*
Analyse given course objectives	The student interprets given course objectives based on prior knowledge and personal values
Summarise results of the group discussion	The student takes an active role in the group discussion and writes a report which represents the outcomes of the discussions
Analyse the study task	The student discusses the study task with the peers and formulates common criteria that the student must meet to carry out the task in a proper way

Table 2.

Effect sizes of main effects for variables concerning the three research questions

variables	Study I	Study II	Study III
Question 1: Development of peer assessment skill			
Using criteria	0.51	1.31	1.43
Constructive comments	ns	1.02	1.22
Structure	ns	0.31	2.61
Naive words	ns	-0.61	ns
Question 2: Improved task performance			
Learning results	ns	0.72	2.24

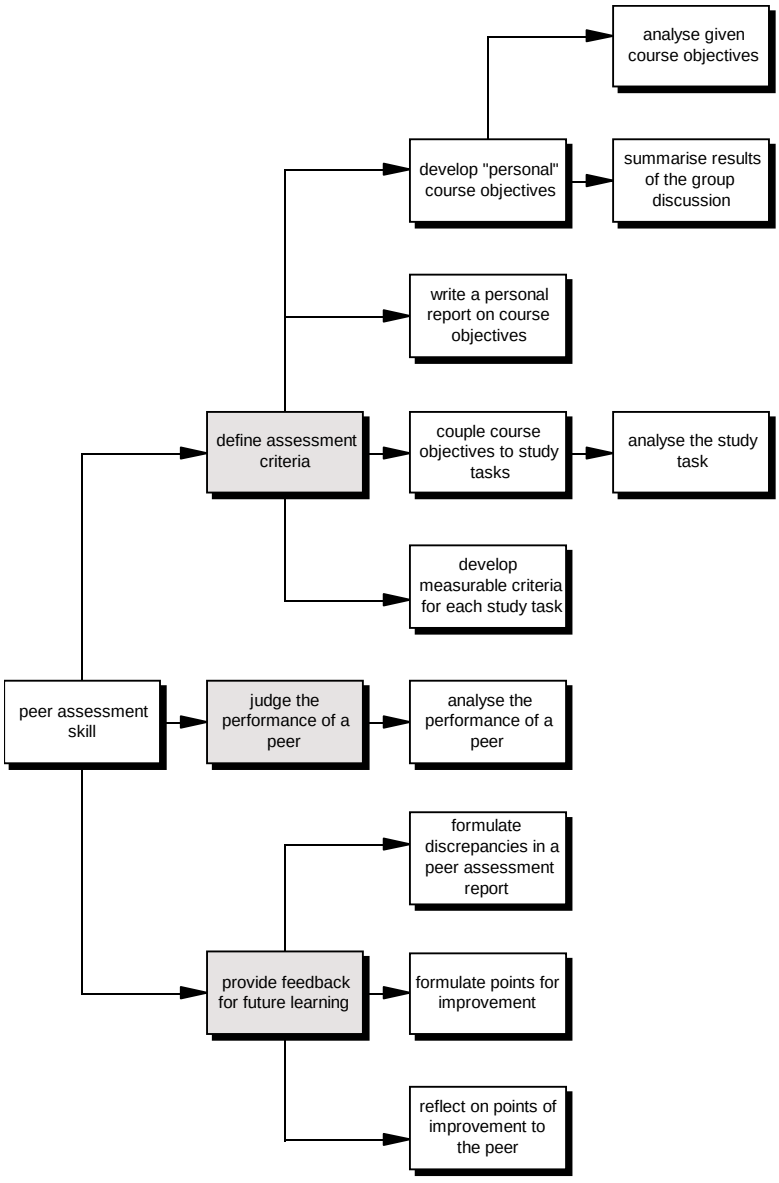
Note: ns = not significant

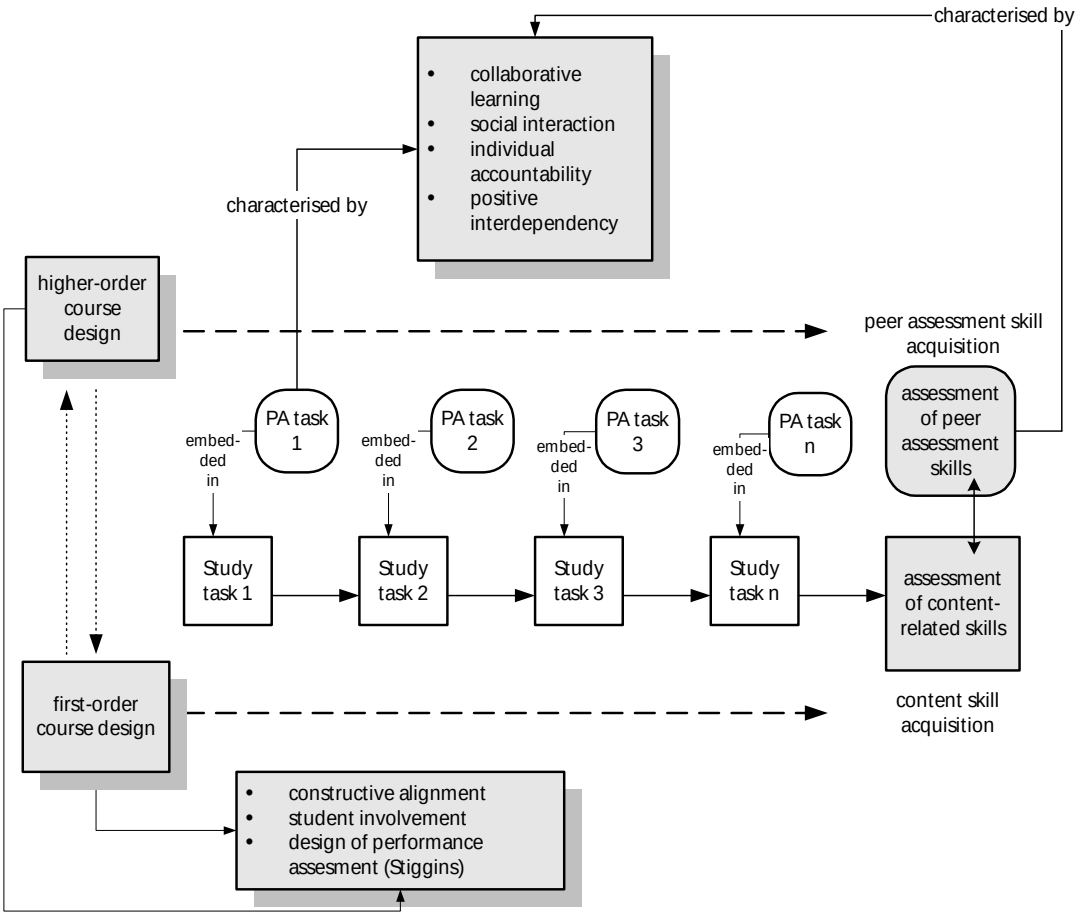
Figure Caption

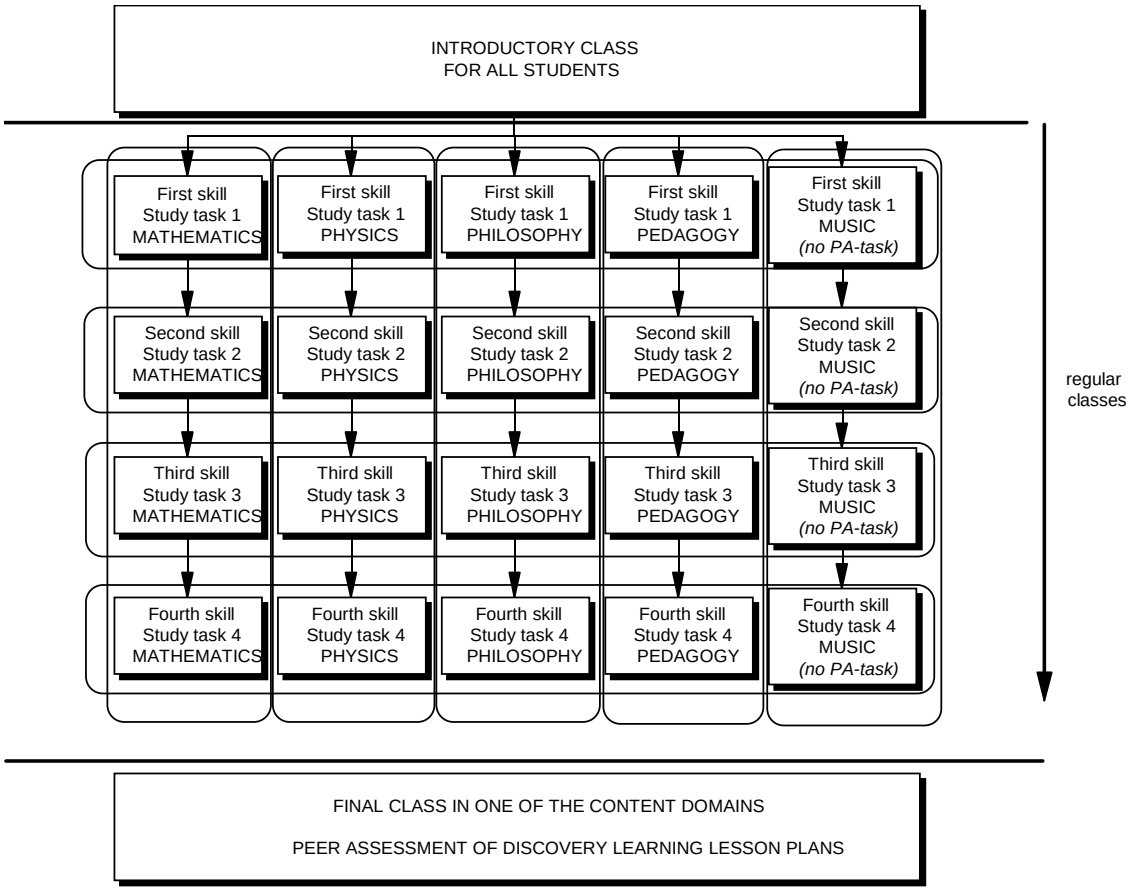
Figure 1. Skill decomposition peer assessment

Figure 2. Student involvement and course design for powerful learning environments – an integrated framework

Figure 3. Organisation of the redesigned course ‘Designing Discovery Learning Lesson Plans’


peer assessment skill
define assessment criteria
judge the performance of a peer
provide feedback for future learning
develop "personal" course objectives
write a personal report on course objectives
develop measurable criteria for each study task
couple course objectives to study tasks
analyse given course objectives
summarise results of the group discussion
analyse the study task
analyse the performance of a peer
formulate discrepancies in a peer assessment report
formulate points for improvement
reflect on points of improvement to the peer


characterised by
collaborative learning
social interaction
individual accountability
positive interdependency
characterised by
higher-order course design
peer assessment skill acquisition
assessment of peer assessment skills
embed-ded in
PA task 1
embed-ded in
PA task 2
embed-ded in
PA task 3
embed-ded in
PA task n
Study task 1
Study task 2
Study task 3
Study task n
assessment of content-related skills
first-order course design
content skill acquisition
constructive alignment
student involvement
design of performance assesment (Stiggins)


INTRODUCTORY CLASS
FOR ALL STUDENTS
First skill
Study task 1
MATHEMATICS
First skill
Study task 1
PHYSICS
First skill
Study task 1
PHILOSOPHY
First skill
Study task 1
PEDAGOGY
First skill
Study task 1
MUSIC
(no PA-task)
Second skill
Study task 2
MATHEMATICS
Second skill
Study task 2
PHYSICS
Second skill
Study task 2
PHILOSOPHY
Second skill
Study task 2
PEDAGOGY
Second skill
Study task 2
MUSIC
(no PA-task)
Third skill
Study task 3
MATHEMATICS
Third skill
Study task 3
PHYSICS
Third skill
Study task 3
PHILOSOPHY
Third skill
Study task 3
PEDAGOGY
Third skill
Study task 3
MUSIC
(no PA-task)
Fourth skill
Study task 4
MATHEMATICS
Fourth skill
Study task 4
PHYSICS
Fourth skill
Study task 4
PHILOSOPHY
Fourth skill
Study task 4
PEDAGOGY
Fourth skill
Study task 4
MUSIC
(no PA-task)
regular
classes
FINAL CLASS IN ONE OF THE CONTENT DOMAINS
PEER ASSESSMENT OF DISCOVERY LEARNING LESSON PLANS